

WHAT IS CLAIMED IS:

1. A computer-implemented method of associating dependency structures from two different languages, wherein the dependency structures comprise nodes organized in a parent/child structure, the computer-implemented method comprising:

associating nodes of the dependency structures to form tentative correspondences; and

aligning nodes of the dependency structures as a function of at least one of eliminating at least one of the tentative correspondences and structural considerations.

2. The computer-implemented method of claim 1 wherein associating includes forming tentative correspondences comprising direct translations.

3. The computer-implemented method of claim 1 wherein associating includes forming tentative correspondences comprising translations of morphological bases and derivations.

4. The computer-implemented method of claim 1 wherein associating includes forming tentative correspondences comprising bases and derived forms of translations.

5. The computer-implemented method of claim 1 wherein associating includes forming tentative correspondences

between nodes wherein one of the nodes comprises more lexical elements than the other node.

6. The computer-implemented method of claim 5 wherein said one of the nodes is a single word in one of the languages and said other node comprises at least two words in the other language.

7. The computer-implemented method of claim 1 wherein aligning pursuant to structural considerations comprises aligning nodes as a function of a set of rules.

8. The computer-implemented method of claim 7 wherein each of the rules of the set of rules are applied to the dependency structures in a selected order.

9. The computer-implemented method of claim 8 wherein each of the dependency structures comprise a set of unaligned nodes and wherein each of the rules are applied successively to the set of unaligned nodes until a set of aligned nodes is identified, then the nodes of the set of aligned nodes are removed from the set of unaligned nodes and each of the rules of the set of rules is again applied successively to the set of unaligned nodes.

10. The computer-implemented method of claim 8 wherein one rule of the set of rules comprises aligning a set

of nodes if a bidirectionally unique translation exists.

11. The computer-implemented method of claim 8 wherein one rule of the set of rules comprises aligning a pair of parent nodes, one from each dependency structure having a tentative correspondence to each other, if each child node of each respective parent node is already aligned to a child of the other parent node.

12. The computer-implemented method of claim 8 wherein one rule of the set of rules comprises aligning a pair of child nodes, one from each dependency structure, if a tentative correspondence exists between them and if a parent node of each respective child node is already aligned to a corresponding parent node of the other child.

13. The computer-implemented method of claim 8 wherein one rule of the set of rules comprises aligning a pair of nodes, one from each dependency structure, if respective parent nodes of the nodes under consideration are aligned with each other and respective child nodes are also aligned with each other.

14. The computer-implemented method of claim 8 wherein one rule of the set of rules comprises aligning a first verb node and an associated child node that is not a verb node from one dependency structure to a second

verb node of the other dependency structure if the associated child node is already aligned with the second verb node, and either the second verb node has no aligned parent nodes, or the first verb node and the second verb node have associated child nodes aligned with each other.

15. The computer-implemented method of claim 8 wherein one rule of the set of rules comprises aligning a pair of nodes, one from each dependency structure, comprising the same part-of-speech, if there are no unaligned sibling nodes, and respective parent nodes are aligned, and linguistic relationships between the set of nodes under consideration and their respective parent nodes are the same.

16. The computer-implemented method of claim 8 wherein one rule of the set of rules comprises aligning a pair of nodes, one from each dependency structure, comprising the same part-of-speech, if respective child nodes are aligned with each other and the linguistic relationship between the set of nodes under consideration and their respective child nodes are the same.

17. The computer-implemented method of claim 8 wherein one rule of the set of rules comprises if an unaligned node of one of the dependency structures having immediate neighbor nodes comprising respective parent nodes, if any, and respective child nodes, if any, all

aligned, and if exactly one of the immediate neighbor nodes is a non-compound word aligned to a node comprising a compound word, then align the unaligned node with the node comprising the compound word.

18. The computer-implemented method of claim 8 wherein one rule of the set of rules comprises aligning a pair of nodes, one from each dependency structure, comprising pronouns if respective parents are aligned with each other and neither of the nodes have unaligned siblings.

19. The computer-implemented method of claim 8 wherein one rule of the set of rules comprises aligning a pair of nodes, one from each dependency structure, comprising nouns if respective parents comprise nouns are aligned with each other and neither of the nodes have unaligned siblings, and where a linguistic relationship between each of the nodes and the respective parents comprise either a modifier relationship or a prepositional relationship.

20. The computer-implemented method of claim 8 wherein one rule of the set of rules comprises aligning a first verb node of one dependency structure to a second verb node of the other dependency structure if the first verb node has no tentative correspondences and a single associated child verb node that is already aligned to the second verb node.

21. The computer-implemented method of claim 8 wherein one rule of the set of rules comprises aligning a first verb node and a single, respective parent node of one dependency structure to a second verb node of the other dependency structure if the first verb node has no tentative correspondences and the single parent verb node is aligned with the second verb node, where the single parent verb node has no unaligned verb child nodes besides the first verb node, and the second verb node has no unaligned verb child nodes.

22. The computer-implemented method of claim 8 wherein one rule of the set of rules comprises aligning a first node comprising a pronoun of one dependency structure to a second node of the other dependency structure, if a parent node of the first node is aligned with the second node and the second node has no unaligned child nodes.

23. The computer-implemented method of claim 8 wherein one rule of the set of rules comprises aligning a first verb node and a respective parent node of one dependency structure to a second verb node of the other dependency structure if the first verb node has no tentative correspondences and the parent verb node is aligned with the second verb node and where the relationship between the first verb node and the parent verb node comprises a modal relationship.

24. A computer-implemented method of associating dependency structures from two different languages, wherein the dependency structures comprise nodes organized in a parent/child structure, the computer-implemented method comprising:

aligning nodes of the dependency structures as a function of a set of rules, the rules being applied to the nodes initially irrespective of the parent/child structure.

25. The computer-implemented method of claim 24 wherein each of the rules of the set of rules are applied to the dependency structures in a selected order.

26. The computer-implemented method of claim 24 wherein later rule applications use an alignment created by an earlier rule application as a reference point that is used to disambiguate between competing alignments.

27. The computer-implemented method of claim 25 wherein each of the dependency structures comprise a set of unaligned nodes and wherein each of the rules are applied successively to the set of unaligned nodes until a set of aligned nodes is identified, then the nodes of the set of aligned nodes are removed from the set of unaligned nodes and each of the rules of the set of rules is again applied successively to the set of unaligned nodes.

28. The computer-implemented method of claim 27 wherein one rule of the set of rules comprises aligning a set of nodes if a bidirectionally unique translation exists.

29. The computer-implemented method of claim 27 wherein one rule of the set of rules comprises aligning a pair of parent nodes, one from each dependency structure having a tentative correspondence to each other, if each child node of each respective parent node is already aligned to a child of the other parent node.

30. The computer-implemented method of claim 27 wherein one rule of the set of rules comprises aligning a pair of child nodes, one from each dependency structure, if a tentative correspondence exists between them and if a parent node of each respective child node is already aligned to a corresponding parent node of the other child.

31. The computer-implemented method of claim 27 wherein one rule of the set of rules comprises aligning a pair of nodes, one from each dependency structure, if respective parent nodes of the nodes under consideration are aligned with each other and respective child nodes are also aligned with each other.



32. The computer-implemented method of claim 27 wherein one rule of the set of rules comprises aligning a first verb node and an associated child node that is not a verb node from one dependency structure to a second verb node of the other dependency structure if the associated child node is already aligned with the second verb node, and either the second verb node has no aligned parent nodes, or the first verb node and the second verb node have associated child nodes aligned with each other.

33. The computer-implemented method of claim 27 wherein one rule of the set of rules comprises aligning a pair of nodes, one from each dependency structure, comprising the same part-of-speech, if there are no unaligned sibling nodes, and respective parent nodes are aligned, and linguistic relationships between the set of nodes under consideration and their respective parent nodes are the same.

34. The computer-implemented method of claim 27 wherein one rule of the set of rules comprises aligning a pair of nodes, one from each dependency structure, comprising the same part-of-speech, if respective child nodes are aligned with each other and the linguistic relationship between the set of nodes under consideration and their respective child nodes are the same.

35. The computer-implemented method of claim 27 wherein one rule of the set of rules comprises if an unaligned node of one of the dependency structures having immediate neighbor nodes comprising respective parent nodes, if any, and respective child nodes, if any, all aligned, and if exactly one of the immediate neighbor nodes is a non-compound word aligned to a node comprising a compound word, then align the unaligned node with the node comprising the compound word.

36. The computer-implemented method of claim 27 wherein one rule of the set of rules comprises aligning a pair of nodes, one from each dependency structure, comprising pronouns if respective parents are aligned with each other and neither of the nodes have unaligned siblings.

37. The computer-implemented method of claim 27 wherein one rule of the set of rules comprises aligning a pair of nodes, one from each dependency structure, comprising nouns if respective parents comprise nouns are aligned with each other and neither of the nodes have unaligned siblings, and where a linguistic relationship between each of the nodes and the respective parents comprise either a modifier relationship or a prepositional relationship.

38. The computer-implemented method of claim 27 wherein one rule of the set of rules comprises aligning a first verb node of one dependency structure to a

second verb node of the other dependency structure if the first verb node has no tentative correspondences and a single associated child verb node that is already aligned to the second verb node.

39. The computer-implemented method of claim 27 wherein one rule of the set of rules comprises aligning a first verb node and a single, respective parent node of one dependency structure to a second verb node of the other dependency structure if the first verb node has no tentative correspondences and the single parent verb node is aligned with the second verb node, where the single parent verb node has no unaligned verb child nodes besides the first verb node, and the second verb node has no unaligned verb child nodes.

40. The computer-implemented method of claim 27 wherein one rule of the set of rules comprises aligning a first node comprising a pronoun of one dependency structure to a second node of the other dependency structure, if a parent node of the first node is aligned with the second node and the second node has no unaligned child nodes.

41. The computer-implemented method of claim 27 wherein one rule of the set of rules comprises aligning a first verb node and a respective parent node of one dependency structure to a second verb node of the other dependency structure if the first verb node has no tentative correspondences and the parent verb node is

aligned with the second verb node and where the relationship between the first verb node and the parent verb node comprises a modal relationship.

42. A computer readable media having information thereon for a computer-implemented machine translation system to translate text from a first language to a second language, the information comprising:

a plurality of mappings, each mapping indicative of associating a dependency structure of the first language with a dependency structure of the second language, wherein at least some of the mappings correspond to dependency structures of the first language having varying context with some common elements, and associated dependency structures of the second language to the dependency structures of the first language also having varying context with some common elements.

43. The computer readable media of claim 42 wherein the dependency structures of said at least some of the mappings have two common elements in each of the languages.

44. The computer readable media of claim 42 wherein the dependency structures of said at least some of the mappings have three common elements in each of the languages.

45. The computer readable media of claim 42 wherein the information includes information indicative of a size of each dependency structure of the first language.

46. The computer readable media of claim 42 wherein the information includes information indicative of an extent of a complete alignment of the dependency structures of the first language originating from a larger dependency structure.

47. The computer readable media of claim 42 wherein the information includes information indicative of a frequency the dependency structure occurred in training data.

48. The computer readable media of claim 42 wherein the information includes information indicative of a type of training data.

49. The computer readable media of claim 42 wherein the information includes information indicative of an extent of the dependency structures originating from a complete parse of the corresponding training data.

50. The computer readable media of claim 42 wherein the information includes information indicative of score related to confidence of alignment of the corresponding dependency structure.

51. The computer readable media of claim 42 wherein at least some of the mappings are indicative of corresponding dependency structures having an element that can vary.

52. The computer readable media of claim 51 wherein the element comprises an under-specified node indicating a part of speech but no specific lemma.

53. The computer readable media of claim 51 wherein the element comprises an under-specified node indicating neither a specified part of speech nor a specific lemma.

54. The computer readable media of claim 51 wherein the element comprises an under-specified node indicating at least one specific syntactic or semantic feature but no specific lemma.